Reducing Credit Union Member Attrition with Predictive Analytics

Nate Derby

Stakana Analytics
Seattle, WA

PhilaSUG 10/29/15

Outline

Introduction: The Problem with Member Attrition

Fisher CU of Ames, IA:

- 6/30/15: 70,359 members, $928MM.
- 3/31/15: 69,534 members, $920MM.
- Grew their membership by 1.19%. That's great, right?
- Wrong! Those 825 new members are a *net* gain.
- We actually gained 3541 *new* members but lost 2716 *existing* members.

Why is this important?

Introduction: The Problem with Member Attrition

Member attrition dramatically affects net membership growth:

- Retained 20% of departing members:
  - $\Rightarrow$ Gained a net 1368 members, 1.97% net membership growth.
- Retained 25% of departing members:
  - $\Rightarrow$ Gained a net 1504 members, 2.16% net membership growth.
- Retained 50% of departing members:
  - $\Rightarrow$ Gained a net 2183 members, 3.14% net membership growth.

Member attrition has a huge effect!

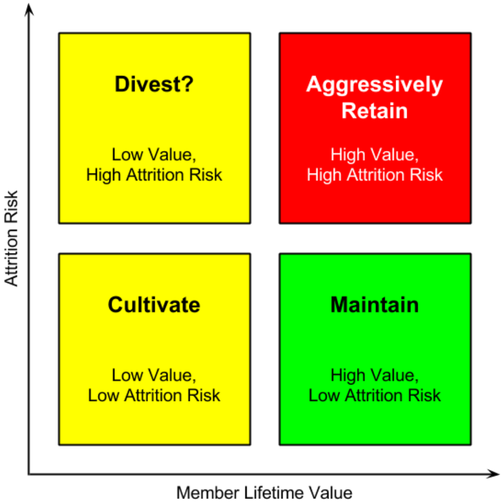
Introduction: The Problem with Member Attrition

Retaining existing members usually easier and less expensive than gaining new members:

- Members already know and trust you, and you already know so much about them.
- Keeping them might be as simple as making a phone call.
- Key is making full use of our member data.

Focus on members with highest value, highest risk of leaving.

Member Segmentation



Many Details to Coordinate!

We're building a *statistical model*:

- Equation that gives the probability that a member will leave.

Data preparation in three steps:

- Duplicating the data.
- Building our variables.
- Partitioning the data.

What Does This Mean?

We're building a statistical model:

$$\text{Probability of leaving in 2-3 months} = f(X_1, X_2, X_3, \ldots)$$

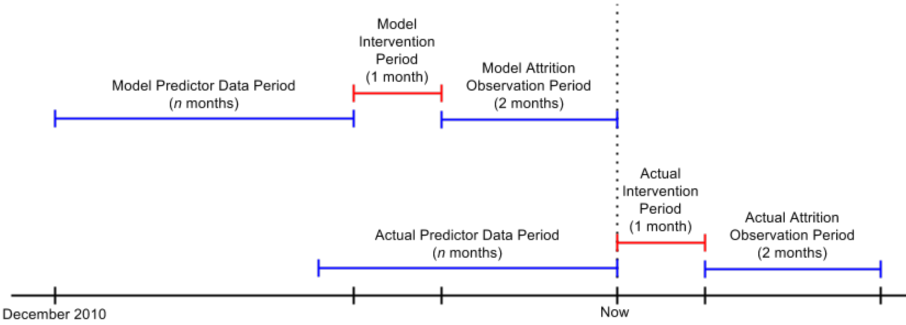
- f is a function we don't know yet (which we'll build).
- Once we know f, we'll use X_1, X_2, X_3, ... to get our probability.

In other words ...

- To *build* the model, we need data as of 3 months ago, coupled with who left 2-3 months later (which we know).
- To *use* the model, we need data as of now, to tell us who are likely to leave 2-3 months later (which we don't know).

The time intervals for X_1, X_2, X_3, ... must be the same for both.

How Do We Do This?



How Do We Do This in SAS?

SAS Code

```
%prepareData( modeling )
%prepareData( scoring )
```

How Do We Do This in SAS?

%prepareData Macro

```
%MACRO prepareData( dataSet );

  %LOCAL now1 now2 now ... attritionEndDate;

  PROC SQL NOPRINT;
    SELECT MAX( effectiveDate )
      INTO :now1
      FROM member_accounts;
    SELECT MIN( tranPostData ), MAX( tranPostDate )
      INTO :startDate, :now2
      FROM member_transactions;
  QUIT;

  %LET now = %SYSFUNC( MIN( &now1, &now2 ) );
```

How Do We Do This in SAS?

%prepareData Macro

```
   %IF &dataSet = modeling %THEN %DO;
       %LET predictorStartDate = &startDate;
       %LET predictorEndDate = %EVAL( &now - 84 );
       %LET attritionStartDate = %EVAL( &now = 56 + 1 );
       %LET attritionEndDate = &now;
       %END;

     %ELSE %IF &dataSet = scoring %THEN %DO;
       %LET predictorStartDate = %EVAL( &startDate + 84 );
       %LET predictorEndDate = &now;
       %END;

   ...

%MEND prepareData;
```

What Are We Doing?

For both of our data sets, we'll build variables that might be predictive of a member closing his/her account.

- We don't care if they're *actually* predictive, as our model will figure that out!
- But we need to “nominate" variables for the model to try out.

Some examples:

- Transaction recency.
- External deposit recency.
- Recent large transaction.
- Small/large number of transactions.
- Seasonality.

How Do We Do This in SAS?

%prepareData Macro

```
PROC SQL NOPRINT;
  CREATE TABLE predictorData1 AS
    SELECT
        id_member,
        MAX( ( &predictorEndDate - tranPostDate )/7 ) AS tranRecency,
        MEAN( ABS( tranAmt ) ) AS meanTranAmt,
        N( tranAmt ) AS nTrans,
        N( tranAmt )/MAX( INTCK( 'month', tranPostDate, &now, 'c' ) ),
      FROM member_transactions
      WHERE
        tranPostDate BETWEEN &predictorStartDate AND &predictorEndDate
          AND UPCASE( tranTypeCode ) IN ( 'CCC', 'CCD', ... 'WTHD' )
      GROUP BY id_member;
```

How Do We Do This in SAS?

%prepareData Macro

```
  CREATE TABLE predictorData2 AS
    SELECT
        id_member,
        MAX( ( &now - tranPostDate )/7 ) AS depRecency,
      FROM member_transactions
      WHERE
        tranPostDate BETWEEN &predictorStartDate AND &predictorEndDate
          AND UPCASE( tranTypeCode ) = 'XDEP'
      GROUP BY id_member;

QUIT;
```

Percentiles via `PROC UNIVARIATE`, then merge ...

Training vs. Validation vs. Test Set

We won't build *one* statistical model for our forecasts.

- We'll build several statistical models.
- We'll choose the one that gives us the best results.
- We'll give an estimate of how accurate those results are.

Each of those steps needs a different data set!

- A statistical model finds the equation that best fits the data.
- If we use the same data, then of course we have great accuracy!
- But the whole point is to predict data we haven't seen yet.

If we never actually tested how well our model predicts unknown data, we could have a nasty surprise.

Training vs. Validation vs. Test Set

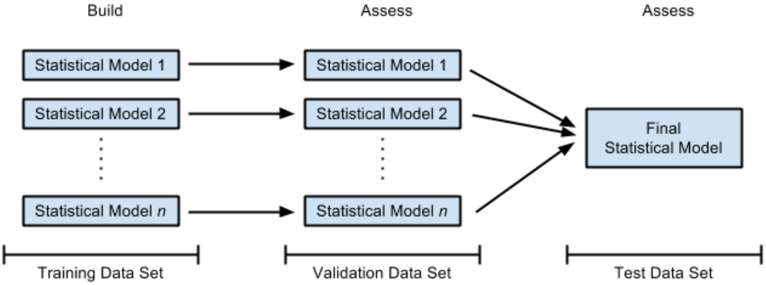
Much better way:

- We'll build several statistical models.
- We'll choose the one that gives us the best results.
- We'll give an estimate of how accurate those results are.

Each of those steps needs a different data set!

- *Training Set* (60%) = data for building the models.
- *Validation Set* (20%) = data for evaluating results of all models.
- *Test Set* (20%) = data for evaluating results of the final model.

Training vs. Validation vs. Test Set



How Do We Do This in SAS?

Training: 60%, Validation: 20%, Test: 40%

```
DATA trainingData validationData testData;
  SET inputData;
  CALL STREAMINIT( 29 );
  randUni = RAND( 'uniform' );
  IF randUni < .6 THEN OUTPUT trainingData;
    ELSE IF randUni < .8 THEN OUTPUT validationData;
    ELSE OUTPUT testData;
RUN;
```

From 69,534 members:

- 41,875 in training set (60.22%).
- 13,807 in validation set (19.86%).
- 13,852 in test set (19.92%).

Building a Model

Building a Statistical Model (with different sets of variables)

```
PROC LOGISTIC DATA=trainingData OUTMODEL=trainingModel1;
  CLASS ageTier( REF='18 and Under' ) / PARAM=ref;
  MODEL attrition( event='1' ) = depRecency ageTier lom
    nProducts calls;
  ODS OUTPUT parameterEstimates = parameter_model1;
RUN;
```

Assessing a Model

Apply the model to our validation set.

Building a Statistical Model (with different sets of variables)

```
PROC LOGISTIC INMODEL=trainingModel1;
  SCORE DATA=validationData
    OUT=validationForecasts OUTROC=validationROC;
RUN;
```

`validationForecasts` and `validationROC` will be used later.

Assessing a Model

If this is our best model and we want to apply it to our test data set:

Building a Statistical Model (with different sets of variables)

```
PROC LOGISTIC INMODEL=trainingModel1;
  SCORE DATA=testData OUT=testForecasts OUTROC=testROC;
RUN;
```

When we're done and want to make our final forecasts:

Building a Statistical Model (with different sets of variables)

```
PROC LOGISTIC INMODEL=trainingModel1;
  SCORE DATA=inputData OUT=finalForecasts;
RUN;
```

Predictive Model

$$\begin{aligned}\text{Attrition Risk} = 1/(1 + \exp(0.07 &- 0.0043X_1 - 2.02X_2 - 3.30X_3 \\ &- 2.91X_4 - 3.21X_5 - 2.76X_6 - 3.37X_7 - 2.36X_8 \\ &- 2.67X_9 - 2.75X_{10} - 4.47X_{11} - 4.17X_{12} + 0.094X_{13} \\ &+ 4.94X_{14} - 2.99X_{15})).\end{aligned}$$

- X_1 = external deposit recency (weeks).
- X_2 = 1 if the member is of age 19-24, 0 otherwise.
- $\vdots$
- X_{12} = 1 if the member is of age 71 or over, 0 otherwise.
- X_{13} = length of membership (months).
- X_{14} = number of products.
- X_{15} = number of customer service calls in the past month.

Interpretation

- For every week that a member goes without an external deposit, his/her odds of attrition multiply by $e^{0.00425} = 1.004$.
- The odds of a member aged 19-24 leaving is $e^{2.0201} = 7.5$ times the odds of a member 18 or under leaving.
- $\vdots$
- The odds of a member aged 71+ leaving is $e^{4.1651} = 64.4$ times the odds of a member 18 or under leaving.
- For every month that a member continues a membership, his/her odds of attrition multiply by $e^{-0.0938} = 0.91$.
- For every product that a member signs up for, his/her odds of attrition multiple by $e^{-4.9355} = 0.0072$.
- For every customer service call that a member makes, his/her odds of attrition multiply by $e^{2.9855} = 19.8$.

Forecasts

Fisher CU Member Attrition Forecasts as of 15-06-30.xlsx - Excel

Member ID	Risk of Attrition in 2-3 months	Aggregate Balance	Age Tier	Length of Membership (Months)	External Deposit Recency (Weeks)	Number of Products	Customer Service Calls (past month)
38050	99.99%	$24,219.38	36 - 40	44	178.01	1	8
37605	99.99%	$2,868.88	36 - 40	6	14.11	1	7
21035	99.99%	$21,795.65	61 - 65	14	28.66	1	7
27143	99.99%	$9,474.69	41 - 45	21	87.23	1	7
8511	99.99%	$202.41	19 - 24	16	53.89	1	7
1006	99.99%	$6,473.96	56 - 60	24	20.07	1	7
32628	99.99%	$13,850.80	66 - 70	22	27.34	1	6
29364	99.99%	$8,423.36	25 - 30	24	96.49	2	8
70252	99.99%	$610.98	19 - 24	32	8.39	1	7
69985	99.99%	$11,398.00	25 - 30	56	197.92	1	7
51167	99.99%	$14,239.38	41 - 45	17	59.12	1	6
68801	99.99%	$105,157.41	25 - 30	22	41.49	1	6
29989	99.99%	$17,112.47	46 - 50	35	140.29	1	6
5638	99.99%	$17,542.99	36 - 40	64	105.08	1	7
23305	99.99%	$7,555.09	46 - 50	30	20.18	1	6
47997	99.99%	$12,247.90	46 - 50	34	91.16	1	6
9277	99.99%	$47,612.51	71 +	8	32.31	1	5
52856	99.99%	$11,555.09	31 - 35	31	126.65	1	6
22334	99.99%	$13,870.93	56 - 60	24	2.84	1	6
25296	99.99%	$5,404.15	31 - 35	34	138.28	1	6

Member Attrition Forecasts

Caveats

This is only half of a retention strategy!

- Nothing will happen without an intervention strategy!

Conclusions

- Predictive analytics can be a powerful tool for member retention
- These techniques just scratch the surface of how we can reduce member attrition with predictive analytics.
- This approach can also be used to predict other aspects of member behavior.
  - When a member will buy a car.
  - When a member will buy a home.
  - Whether a member is committing a financial crime.
- We can also use these techniques to further cultivate our members.

Further Resources

- Junxiang Lu.
  Predicting Customer Churn in the Telecommunications Industry.
  Proceedings of the Twenty-Seventh SUGI Conference, 2002.
- Ward Thomas.
  Improving Retention by Predicting Both Who and Why.
  Proceedings of the Twenty-Third NESUG Conference, 2010.

Nate Derby: `nderby@stakana.com`